

| | |
|---------------|--|
| Title | Speech and Impression : A Study of Recognition of Emotion by Japanese, Chinese and English |
| Author(s) | Taniguchi, Miki |
| Citation | 大阪大学言語文化学. 12 p.237-p.252 |
| Issue Date | 2003-03-31 |
| oaire:version | VoR |
| URL | https://hdl.handle.net/11094/77961 |
| rights | |
| Note | |

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

Speech and Impression: A Study of Recognition of Eemotion by Japanese, Chinese and English*

TANIGUCHI Miki**

Keywords : prosody, paralinguistic information, impression

近年、音声研究の分野では、パラ言語情報を扱った研究が盛んになってきている。藤崎（1994）は、音声が伝達する情報を(1)言語記号によって伝えられる「言語情報」(2)対話の制御や発話の意図、話者の感情などに関する「パラ言語情報」(3)性別や個人的な特徴などを担う「非言語情報」の3つに分類することを提案している。本稿では、藤崎の分類による(2)のパラ言語情報の伝達、特に韻律と印象という点に注目し、合成音を用いて、日本語、中国語、英語を母語とする話者に印象評価実験を行い、パラ言語の普遍性についての一考察を行った。

実験には、計49名（日本語母語話者16名、中国語母語話者17名、英語母語話者16名）が参加し、58の音声刺激に対して「肯定的な印象」から「否定的な印象」までを7段階評価で回答した。

実験結果から、発話の長さは少なからず聞き手の印象評価に影響を与えている可能性が指摘でき、実験における最も短い発話と長い発話においては、3言語話者ともにやや否定的な印象を受けていることが分かった。一方、ピッチの高さの変化に関しては、印象評価に有意な差がみられなかった。この結果は、Van Bezooijen et al. (1983)をはじめ、特定の音声について認識テストを行った先行研究の結果と比較する上で有意義なものであると思われ、特定の韻律（発話全体の長さやピッチの高さ）に関して、音声と感情評価における普遍性と非普遍性を指摘する資料を与えているといえる。

1 Introduction

When considering human communication, it is obvious that we do not communicate with each other only by the literal meaning of our utterances. Even though the same sentence might be produced, a different meaning can be conveyed in a different situation. Some of the main factors

* 発話と印象－日本語、中国語、英語母語話者を対象にした感情の認識－（谷口未希）

** 大阪大学大学院言語文化研究科博士後期課程

giving a different meaning to the same sentence are paralinguistic information such as pitch, loudness, tempo, pause and so on. The correlation between prosody and intention or emotion, which is one topic of studies in phonetics and psycholinguistics, has been researched over many years. However, there are still unexplored issues in this area.

The purpose of this paper is to investigate whether or not vocal impressions are universal. In it we focus particularly on the effects of utterance duration and pitch level. Previous studies characterized prosodic features for specific emotions, mostly basic emotions such as happiness, sadness, fear, anger and disgust (Fairbanks and Pronovost 1938, Williams and Stevens 1972, Frick 1986, Scherer 1981, 1986, 1995, Takagi and Umeda 1989, Maekawa 1997 among others). Those studies suggested that anger, happiness, confidence, and fear are signaled by increased pitch, increased loudness, and a fast rate of speech; and that contempt, boredom, and grief have low pitch and slow rate, though these correlations do not always correspond in all studies. On the other hand, from the cross-cultural viewpoint, Kramer (1964) stated that there were few differences between American and Japanese in the expressions of emotion. Chang (1958) concluded that Chinese do not have much freedom in using pitch to express emotion. Sogon (1975) found that Japanese speakers recognized grief most readily, but they were less able to recognize contempt. Van Bezooijen et al. (1983) examined the recognition of emotion in Dutch by Dutch, Taiwanese and Japanese listeners, and concluded that all three-language subjects could identify emotion correctly. Another finding was that Taiwanese and Japanese most accurately identified sadness and anger, while joy was more problematic.

To test the universality of prosodic features associated with emotion and mood, which previous studies reported as the universal, I did a perception test of a Japanese sentence by speakers of three different language: Japanese, Chinese and English. If the vocal impression is universal, the judgments by Chinese and English people should be similar. In addition, those results should also indicate a similar trend to the results with Japanese.

Having taken the definitions mentioned above into account, this paper has adopted the definitions of mood and emotion by Murray and Arnott (1993), which were based on Oatley (1989).

Emotions are different from moods --- emotions arise suddenly in response to particular stimuli and last for seconds or minutes, while moods are more vague in nature, lasting for hours or days; emotions can broadly be considered as being concerned with changing something, and moods with maintaining something.

(Murray and Arnott 1993: 1097)

We have also used *positive* and *negative* impression as regards the listener's evaluation. According to Schlosberg (1954), positive-negative corresponds to 'valence', which is one of three axes used to explain emotion by a three-dimension model. In that model, valence can range from love and happiness to anger. Although our definition is based upon that by Schosberg, in this paper positive-negative means not only pleasantness-unpleasantness, but also includes wider and opposite feelings such as happy-unhappy, delight- non delight, satisfied-unsatisfied, excited-unexcited and so on.

The next section will describe the design of the experiment. The results are discussed in section 3. The final section states the conclusions.

2 Method of the Experiment

2.1 Subjects

An experiment was conducted with native speakers of three different languages, Chinese, English and Japanese. 49 people, 16 native Japanese speakers (10 males and 6 females), 17 native Chinese speakers (8 males and 9 females) and 16 native English speakers (8 males and 8 females) participated in the experiment. They were aged between their 20's and 30's. None of the participants has a problem in terms of their sense of hearing. None of the Chinese or English subjects had studied Japanese before this experiment, nor knew any Japanese at that time.

2.2 Materials

The target token was IINE in Japanese, meaning something similar to 'It is good, isn't it' in English. An original target stimulus was recorded by an adult, male, Tokyo Japanese dialect speaker. Regarding this token, the speaker's mood/intention corresponded with the native Japanese listeners' impression in the perception test¹⁾. As stimuli in the experiment, this original token was synthesized in terms of duration and pitch. The characteristics of the original data were that the duration of the token was 548ms, minimal F0 was 78.7Hz and maximum was 128.7Hz. Using 'Praat' ²⁾ the duration of the original token was changed from half-length to double-length that of the original token, by 10% steps. In other words, the original utterance was as 100%, the half-length was 50%, and 10% steps via 60% length, 70% length, 80% length were

¹⁾The result of a perception test is unpublished. The subjects numbered 24, aged 20's to 30's. Subjects listened to the token and carry a force-judgment task: a positive feeling, a negative feeling and either.

²⁾Praat which was developed by Dr. Paul Boersma and Dr. David Weenink, Institute of Sciences in University of Amsterdam, was used for synthesis. URL: [thhp://www.praat.org](http://www.praat.org)

added to reach 200% length as shown in Figure 1. Thus, the number of stimuli was 16, including the original token. On the other hand, for fundamental frequency the original token was changed each plus-minus one semitone step, and the maximum was plus-minus seven semitones as in Figure 2. The number of stimuli was 15 in terms of pitch, including the original token.

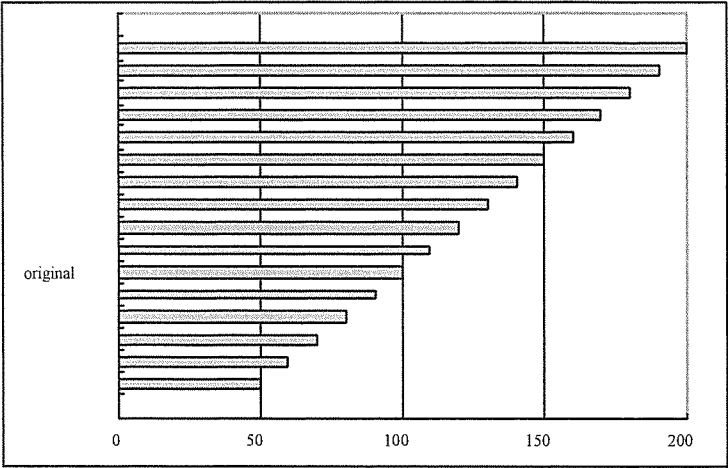


Figure 1 The method of manipulation for duration.
The horizontal axis shows rate of extending (%)

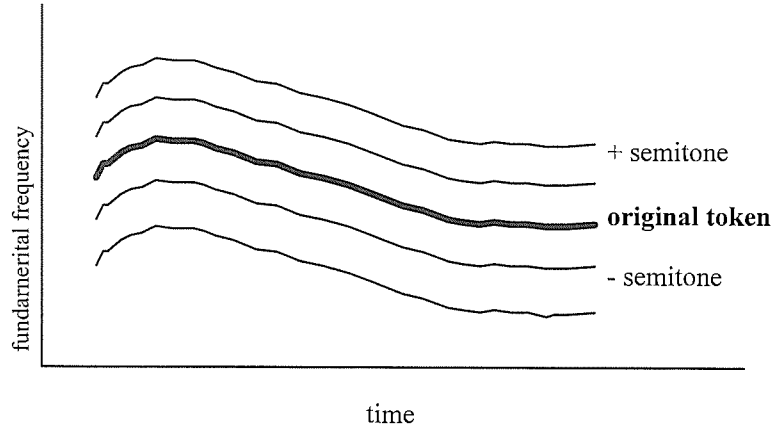


Figure 2 The method of manipulation for pitch level.
The horizontal axis shows time, and the vertical axis shows F0.

In addition to synthesized stimuli, three kinds of anchor stimuli were used in order to test the correspondence of listeners' evaluation during the experiment. Three different Japanese

speakers, one male and two females produced these anchor tokens. The number of moras and phonemes were the same as the original token above. The experiment consisted of 16 stimuli each of which had different length, 15 stimuli each of which had different pitch level, and also contained 27 anchor tokens. Thus the number of stimuli was 58 in total.

2.3 Procedure

Before the experiment, the subjects were told that the aim of the experiment was to investigate the impressions created by different tokens. At the beginning the test they also wrote on the paper some information about themselves such as sex, age, language background (for Chinese and English) and academic background.

The subjects listened to the target tokens one by one, from the computer and then were asked to judge their impression for each token. They evaluated each token using seven-point scales, to which were added adjective descriptors such as 'very negative,' 'negative,' 'a little negative,' 'uncertain,' 'a little positive,' 'positive,' and 'very positive.' All answers were written on the question sheet. The definition of 'positive' in the experiment was given as something that was like happy, glad, pleased, delighted, excited, interested, relaxed, calm, content, satisfied. The definition of 'negative' was given as something that was like sad, depressed, gloomy, bored, distressed, annoyed, angry, afraid. These adjective words were based on Russell (1980). Additionally, the experimenter did a pilot study before, in order to check the correlation between 'positive / negative' and those adjective words through the same sounds.

The three practical trials were prepared before the test. In the test section, a stimulus was given twice for each question, and an interval of about 5 seconds was provided between the questions.

3 Results and Discussion

3.1 Duration

As for restrictions, the data from 5 subjects (two Chinese males, a Chinese female, an English male and a Japanese male) were excluded because their responses did not obviously correspond with each other for the same anchor stimuli. Therefore, 44 subjects' data were analyzed.

3.1.1 Japanese listeners

Figures 3 to 6 show the distribution of evaluation by Japanese for tokens of 70%, 100%,

130% and 180% duration. Figure 3 indicates that judgment of the 70% duration utterance had a wide range, from 2 to 6 of evaluation value. As evaluation 4 means 'uncertain', some subjects judged the sound as negative, and others judged it as positive. In contrast, Figure 4 shows that the original utterance, that is the 100% duration utterance, had a range from 4 to 6. This means that 100% duration stimuli made a positive impression, though it had uncertain judgment; but at least it was not a negative one. Comparing Figure 3 with Figure 4, we can see that the distribution changes from ambiguous judgments to positive ones. Figure 5 also shows a positive trend as a whole, though it had a range from 3 to 6. According to Figures 5 and 6, the bar indicating number 4 in the horizontal axis became shorter from Figure 5 to Figure 6. Subjects who judged as uncertain for the 100% duration token would be likely to have a positive impression for the 130% duration token. Finally, in Figure 6, the tallest bar indicated the evaluation 3, which means a little negative. Although some subjects chose 4 or 5, which mean uncertain or only a little positive, more than half of the subjects clearly judged it as a negative impression. Therefore, as Figures 5 and 6 show, the listeners' impression has changed from positive to negative, again. Through Figures 3 to 6, we can see the distribution shifting slightly from left to right, and right to left again. This means that the listeners' impression also has shifted from being a negative to a positive impression, and back to a negative impression; depending on duration of the whole utterance. Figure 7, which was based on all data like Figures 3 to 6 and similar figures for other duration, presents the median of evaluation. This Figure clearly shows the shifting between a negative and a positive impression which we described above. That is, the evaluation of the token's impression shows a trend that when the tokens are shorter, the impression is uncertain or at least less positive, and when the duration of the token is longer, the evaluation is also less positive.

We offer here some reasons why the shorter token made a less positive impression and the longer token showed a slightly negative impression.

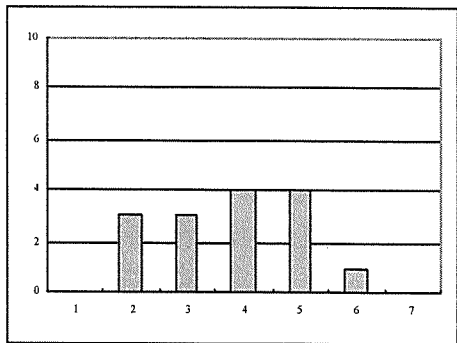


Figure 3 The distribution of evaluation with 70% duration token by Japanese.
The vertical axis shows the number of subjects.
The horizontal axis shows value of evaluation.

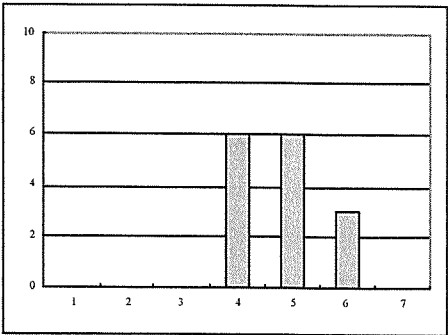


Figure 4 The distribution of evaluation with 100% duration token by Japanese

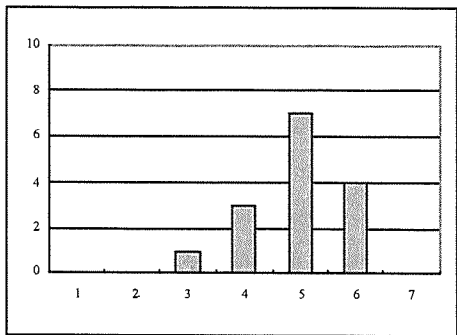


Figure 5 The distribution of evaluation with 130% duration token by Japanese

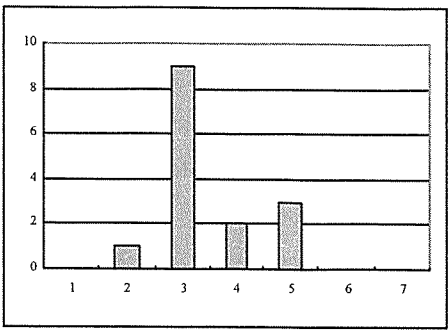


Figure 6 The distribution of evaluation with 180% duration token by Japanese

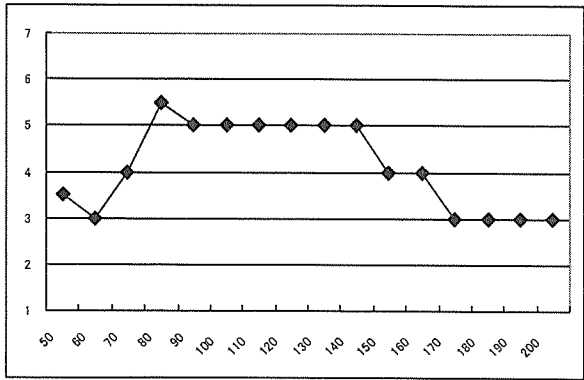


Figure 7 The median of Japanese listener's evaluation with duration for all stimuli. The vertical axis shows the number of subjects. The horizontal axis shows the kinds of stimuli from 50% duration to 200% duration.

One is associated with the results of the production test. Taniguchi (2002) reported that some Japanese native speakers produced the same token with a shorter duration in a negative situation, than in a positive situation. That is, it is natural in a sense that an utterance produced with a negative mood is short, and it is judged as negative by a listener. Secondly, a sentence-final particle 'ne' is an important cue, not only to make an impression but also to serve a pragmatic or semantic function (Takubo 1990, Katagiri 1995, Takubo and Kinsui 1996). Takubo and Kinsui (1997) stated that the sentence-final particle was characterized as a marker for operations being carried out on the database, and in particular 'ne' was a marker when the speaker was in the processing of incorporating assumptions from the indirect-domain into the direct domain. Having considered the acoustic information, if the sentence-final particle 'ne' is produced with a very short duration (except in the case of confirmation), the utterance must be thought of as impolite in Japanese. In this experiment, when the whole duration was short, duration of 'ne' was of necessity even more short. Thus, the utterance as a whole could cause an impolite impression, or negative impression. On the other hand, thinking of the reasons why the utterance of very long duration also gave a negative impression, when the whole duration of the token is long, 'ne' duration also is long. Where the sentence-final particle 'ne' is produced with a very long duration, it should indicate that listener does not have sympathy with the speaker. In Japanese, the sentence-final particle is one of the important keys for communication. Changing the whole utterance duration means changing 'ne' duration, and this could make a negative impression.

3. 1. 2 Chinese and English listeners

Figure 8 presents the median of evaluation for Chinese listeners. Similarly, Figure 9 presents the median of evaluation for English listeners. These data are analyzed similarly to the data in Figures 3 to 6. Figure 8 tells us that shorter tokens, 60% to 100% length, carry more positive impression and longer tokens, more than 110% duration, carry less positive impression, though the tokens are still keeping their positive evaluation. We can also see from Figure 8 that the shortest and longest tokens were judged as giving negative impression.

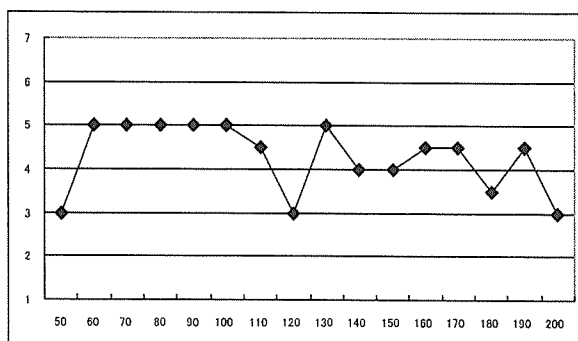


Figure 8 The median of Chinese listener's evaluation with duration for all stimuli.

On the other hand, examining the median of English in Figure 9, there is much less movement associated with duration. Evaluation always indicates 3 or 4. This means English subjects had a similar impression for most stimuli, and the impression was generally neutral rather than being a strong impression. Again, what is clear here is that the 50% duration token is identified as being a little negative. In addition, judgments are often seen as negative when the duration of a token becomes longer and longer.

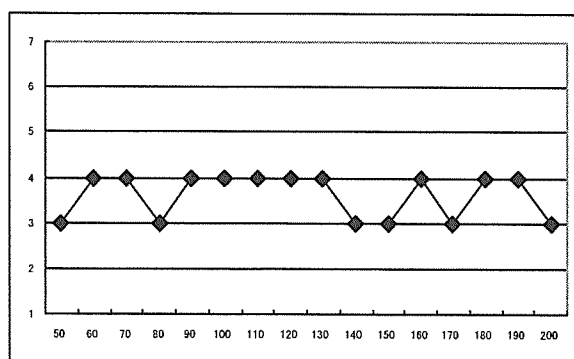


Figure 9 The median of English listeners' evaluation with duration for all stimuli.

Considering the difference between Chinese and English results, one was that the Chinese indicated a trend related to the token duration, which the English did not have. That is, where the token is longer, the evaluation of that token is less positive in Chinese. This is a similar trend to

Japanese, though it is a less clear distribution than that for Japanese. In addition, the range which the Chinese used for evaluation was similar to Japanese judgment. However, there is a point which cannot be explained, the dip in 120% in Figure 8. We could not find why 120% duration made a negative impression, in spite of two points besides 120% being judged as positive. One might be the order of stimuli. Another difference between the Japanese graph and the Chinese one is distribution between 180% to 200%. This may show that the Chinese still experienced a little confusion. The other finding is that the Chinese impression was more positive than that of the English, as a whole. English subjects judged tokens as neither negative nor positive, or a little negative. Most of the English listeners did not gain a positive impression from the target tokens. This might be because the utterance meaning basically depends on the intonation system in English, such as falling, rising and those in combination (Cruttenden 1986). English listeners, therefore, might have used pitch movement as a cue to judge the stimuli, rather than duration. However, as our stimuli had exactly the same intonation pattern, their judgment could not change so much.

In consequence, Japanese and Chinese listeners judged longer tokens as being more negative. In addition, they also had a negative impression of the shortest token given by those two language speakers. This means impressions of the token have a strong connection with the token duration, for Japanese and Chinese listeners. A short token seems to sound impolite, and a long token does not sound interesting. However, English listeners did not change their evaluation much for the different duration, probably because of the English intonation system. The duration may more or less influence the impression received by the listener for some languages, but our data did not clearly show the universality as a whole in terms of duration.

3. 2 Pitch

Regarding restrictions, the data from 3 subjects (two Chinese females and an English male) were excluded because their responses did not obviously correspond with each other for the same anchor stimuli. Therefore, 46 subjects' data were analyzed in total.

3. 2. 1 Japanese Listeners

Figure 10 shows the median evaluation for the different pitch tokens by Japanese listeners. We see from Figure 10 that there is no correlation between pitch level and the evaluation of impression, because evaluation values are always between 4 and 5, and the distribution concerning the difference of duration looks disordered. There is no significant correlate ($r=.011$, $P=.05$). This means that a difference of pitch level does not give a different

impression. This too, does not correspond with the results of pitch average and pitch range in previous studies. Levin and Lord (1975 in Murray and Arnott 1993) reported that pitch was an important emotional factor. Williams and Stevens (1972) also stated that relative pitch was a better cue for the expression of emotion than absolute pitch.

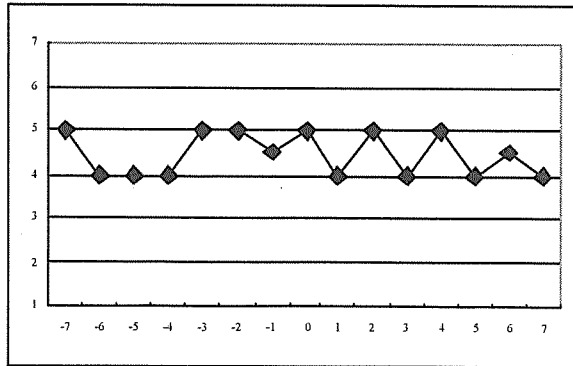


Figure 10 The median of Japanese listener's evaluation with pitch level for all stimuli.

This result was also different from what we expected. We anticipated that higher pitch would give a more positive impression, and lower pitch a more negative impression. One possibility causing disagreement could be the method of manipulation. The stimuli of the experiment were manipulated with only pitch level, not pitch range. This means that pitch contour and pitch range were the same as the original utterance, but only the fundamental frequency was changed by plus-minus one semitone step each, as Figure 2 shows in the previous section. If the pitch contour affects the vocal impression, changing only the fundamental frequency in other words, pitch level would not make sense. Since the original token was produced with a positive mood by the speaker in this case, even though the pitch may be changed to any level, listeners might judge them as giving the same impression. That is because the pitch range itself is not changed. In addition to the pitch level, timbre in all stimuli is not changed, either. The stimuli are not characterized by nature such as keen, soft and so on. Thus, impression for any stimuli by listeners did not change in this experiment. Williams and Stevens (1972) commented about pitch of an utterance in relation to the speaker's normal pitch range. However, it seems that identifying the relative pitch of speakers was difficult in this case. It is also not certain whether the subjects noticed that some tokens were based on the same speaker.

To sum up, at least the difference of pitch level did not affect the listeners' impression in respect of positive-negative evaluation. The impression which the token gives might not change,

if the pitch contour is the same. Other possible factors associated with the different impression would be pitch range, pitch movement or timbre, rather than pitch level.

3. 2. 2 Chinese and English Listeners

Figure 11 shows the median of evaluation with Chinese listeners, and Figure 12 with English listeners. They indicate a slightly different trend. The Chinese result, in Figure 11, shows clearly that high pitch gives a more positive impression. There is significant correlation ($r=.83$, $P=.001$). Most judgments show as positive. On the other hand, the English data show more ambiguity ($r=-.09$), but, at least, their judgments indicate the positive. Interestingly, the Chinese data shows a wide range of evaluation associated with different stimuli, while by contrast the English data have a narrow range in terms of evaluation. This comparison is the same as the duration result between Chinese and English. The Chinese data range is bigger than that of English.

To compare with the Japanese results, the English result is similar to the Japanese one, because Japanese evaluation also did not indicate a correlation between pitch level and evaluation. Therefore, only Chinese listeners judged impressions depending on pitch levels. Since it is occasionally suggested that many variations of Chinese use a high register, the result here might be related to that. However, there is still potential for investigating this in more detail.

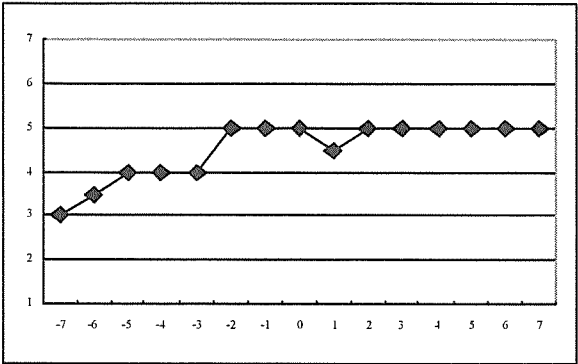


Figure 11 The median of Chinese listeners' evaluation with pitch level for all stimuli.

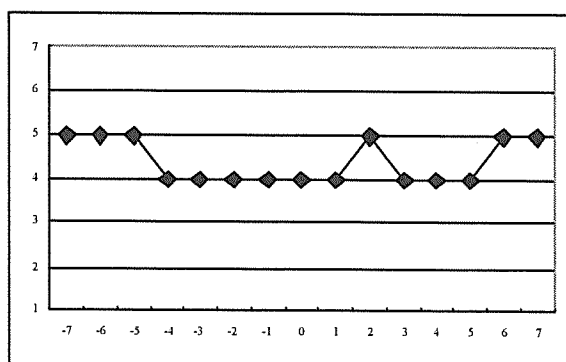


Figure 12 The median of English listeners' evaluation with pitch level for all stimuli.

4 Conclusion

We tested the effect on duration and pitch associated with the listener's impression of a Japanese token by native and non-native speakers: Japanese, Chinese and English native speakers. We found that when the token duration is longer, the impression gained is less positive, and also that a shorter token would sometimes give a negative impression, for the Japanese. The Chinese trend was very similar to Japanese. On the other hand, the English trend was slightly different from the other two, that is to say their judgment was not strongly affected by duration.

The other finding was that there was not a strong connection between the listener's impression and different pitch levels for Japanese and English, but there was some relation for Chinese, though it was not significant. Compared with the effect of duration and pitch, duration has more effect upon the listener's impression than pitch, because listeners' evaluations indicated positive rather than negative impressions for most cases.

Van Bezooijen et al (1983) suggested that Taiwanese and Japanese people, who were very limited in terms of any contact with speakers of Dutch, could correctly identify Dutch vocal expression of emotion beyond chance expectancy. Their results also suggested that there may be characteristics of the emotional voice that are universally recognizable. Our results, however, would not seem to show such universality in terms of vocal emotion, although there is a minor connection in that the shortest and the longest token were judged as negative by all three language speakers. Another finding in Van Bezooijen et al (1983) was that sadness and anger were identified most accurately, and joy was not. This corresponded with the findings for the cross-cultural recognition of facially expressed emotion (Boucher and Carlson 1980). The results

of the present study are slightly different on this point. Subjects judged tokens as positive more than negative for Japanese and Chinese listeners, but this result was not clear for English listeners. One reason why most listeners made a positive rather than negative judgment in this experiment might be that the original token which was synthesized, was produced with a positive feeling. This suggests that there might be another stronger cue than utterance duration by which listeners judge the speaker's mood. (For more details see Taniguchi 2002).

In conclusion, there was no strong evidence to support the universality associated with vocal impression and mood. However, it suggested that longer tokens make a negative impression, and the shortest tokens also give a negative impression and this trend would seem to correspond for at least some languages. People who do not know the language can identify the mood of a speaker correctly, or at least more or less correctly. The utterance duration should be one of the cues used to recognize the mood and emotion.

References

- Chang, N.-C. T. (1958) Tone and intonation in the Chengtu dialect. *Phonetica* 2: 59.
- Cruttenden, A. (1986) Intonation. New York: Cambridge University Press.
- Fairbanks, G. (1940) Recent experimental investigations of vocal pitch in speech. *Journal of the Acoustical Society of America* 11: 457-466.
- Fairbanks, G. and W. Pronovost. (1938) Vocal pitch during simulated emotion. *Science* 88/2286: 382-383.
- Fairbanks, G. and W. Pronovost. (1939) An experimental study of the pitch characteristics of the voice during the expression of emotion. *Speech Monograph* 8: 85-91.
- Frick, R.W. (1985) Communicating Emotion: The Role of Prosodic Features. *Psychological Bulletin* 97: 412-429.
- Fujisaki, O. (1994) Inritsu kenkyu no sysokumen to sono kadai [Current and further issues in research of prosody.] Nihon onkyou gakkai kouen ronbunshu [In Proceedings of the Acoustic Society of Japan], 287-290.
- Katagiri, Y. (1995) Syujyoshi niyoru taiwa tyousei [Discourse management by the sentence-final particle]. *Gengo [Language]* 24-11: 38-45.
- Kramer, E. (1964) Elimination of verbal cues in judgments of emotion from voice. *Journal of Abnormal and Social Psychology*, 68: 390-396.
- Maekawa, K. (1997) Onsei ni yoru jyouhou dentatsu no mekanizumu [communication mechanism with voice]. *Nihongogaku [Study of Japanese]* 16: 95-105.

- Murray, I. R. and John L. Arnott (1993) Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*. 93. 1097-1108.
- Pollack, I., H. Rubenstein, and A. Horowitz,. (1960) Communication of verbal modes of expression. *Language and Speech* 3: 121-130.
- Russel, J.A. (1980) A Circumplex Model of Affect. *Journal of Pwesonalityand Social Psychology* 39: 1161-1178.
- Scherer, K.R. (1981) Speech and Emotion States. In Darby, J.K. (ed.) *Speech Evaluation in Psychiatry*. New York: Grun and Srratton
- Scherer, K.R. (1986) Vocal Affect Expression: A review and a model for future research. *Psychological Bulletin* 99: 143-165.
- Scherer, K.R. (1995) Expression of Emotion in Voice and Music. *Journal of Voice* 9: 235-248.
- Sogo, S. (1975) A study of the personality factor which affects the judgment of vocally expressed emotions. *Japanese Journal of Psychology*, 46: 247-254.
- Sugito, M. (2001) Jyosi 'ne' no imi kinou to intonasyon [meaning, function and intonation in the sentence final particle 'ne' in Japanese.] In onsei bunpou kenkyukai [Spoken Language Working Group] (ed.) *Bunpou to onsei III [Speech and Grammar III]*: 3-16.
- Takagi, T. and H.Umeda (1989) Koeshitsu henkan ni yoru kanjyou hyougen seigyō no tameno butsuriteki youin no kaiseki [Analysis of acoustic feature for control of emotional expressions of speech by voice-quality conversion]. Proceedings of the Acoustic Society in Japan 3-3-4: 275.
- Takubo, Y. (1990) Danwa kanri no riron [Theory of discourse management]. *Gengo [Language]* 19-4: 52-58.
- Takubo, Y, and S. Kinsui. (1996) Taiw to kyōyū chishiki [Discourse and mutual knowledge]. *Gengo [Language]* 25-1: 30-39.
- Taniguchi, M. (2002) An experimental study of speech and recognition related to mood in Japanese. A thesis for the MA in Language, Mind and Communication at University of Newcastle upon Tyne.
- Tickle, A. (2000) English and Japanese speakers' emotion vocalization and recognition: a comparison highlighting vowel quality. Proceedings of ISCA workshop. Belfast, Ireland.
- Utsuki, S. (1979) Onsei ni yoru kanjyou hyōsyūto ninti. [emotion and recognition.] Proceedings of the Psychological Society in Japan: 43. 387.
- Van Bezooijen, R., S.A. Otto, and T.A. Heenan (1983) Recognition of vocal expressions of

emotion: A three-nation study to identify universal characteristics. *Journal of Cross-Cultural Psychology* 14: 387-406.

Williams, C.E. and K.N. Stevens (1972) Emotions and Speech: Some Acoustical Correlates. *Journal of the Acoustical Society of America*. 52: 1238-1250.

Yanagida and Takizawa, M. (1994) hiniku hyougen ni okeru inritsuteki tokutyo [Prosodic features of sarcasm.] *Nihon onkyou gakkai kouen ronbunshu* [In Proceedings of Acoustic Society of Japan]:389-390.